Journal Title
XX(X):1–22

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Analysis of Urban Vibrancy and Safety in Philadelphia

Colman Humphrey[1] and Shane T. Jensen[1] and Dylan Small[1] and Rachel Thurston[2]

Abstract

Statistical analyses of urban environments have been recently improved through publicly available high resolution data and mapping technologies that have been adopted across industries. These technologies allow us to create metrics to empirically investigate urban design principles of the past half-century. Philadelphia is an interesting case study for this work, with its rapid urban development and population increase in the last decade. We outline a data analysis pipeline for exploring the association between safety and local neighborhood features such as population, economic health and the built environment. In particular, we focus on quantitative measures of the built environment that serve as proxies for *vibrancy*: the amount of human activity in a local area. Historically, vibrancy has been very challenging to measure empirically. Measures based on land use zoning are not an adequate description of local vibrancy and so we construct a database and set of measures of business activity in each neighborhood. We employ several matching analyses within census block groups to explore the relationship between neighborhood vibrancy and safety. We find that neighborhoods with more vacancy have higher crime but within neighborhoods, crimes tend not to be located near vacant properties. We also find that more crimes occur near business locations but businesses that are active (open) for longer periods are associated with fewer crimes.

Keywords

urbanism; vibrancy; crime; matching;

[1]Department of Statistics, The Wharton School, University of Pennsylvania
[2]Stantec

Corresponding author:

Shane T. Jensen, Department of Statistics The Wharton School, University of Pennsylvania, 463 Huntsman Hall 3730 Walnut Street, Philadelphia, PA, USA 19102
Email: stjensen@wharton.upenn.edu

1 Introduction

Throughout history there have been many perspectives on the approach to planning of cities, with a notable clash between dense, organically-formed urban spaces versus large-scale clearing and planning of "superblocks" and automobile-centric layouts. The former perspective viewed city development as a social enterprise created by many hands, whereas the top-down central planning approach involved less input from the residents affected by city changes. The urban renewal movement of the 1960s and 1970s is the largest example of this latter effort, but the same mentality still drives many current development decisions.

One historical motivation for top-down urban renewal projects was the idea that cities were over-crowded. Winsborough (1965) discusses both positive and negative perspectives on the effects of population density in urban settings. Population density has been positively associated with division of labor but has also been linked to psychological strain and negative health outcomes. Simmel (1971) argues that the emotional stress caused by high population density produces negative attitudes and hostility among the populace. In a study of Baltimore, Verbrugge and Taylor (1980) find both positive and negative effects of population density and suggest that population size is a more important factor for attitudes and behavior in urban environments.

Earlier responses to anti-density rhetoric and the challenges of urban living during the industrial age resulted in code regulations, restrictive land use zoning, and sometimes large scale clearing of entire neighborhoods. During the age of urban renewal, dense urban environments were demolished and replaced by trending architectural works, civic monuments and tree lined boulevards built for reducing population density and easing automobile traffic, along with large housing projects for displaced communities. Over time, a large number of these projects failed to attract pedestrian activity and resulted in high crime housing areas.

In her seminal work *The Death and Life of Great American Cities* (1961), Jane Jacobs challenged the analyses of proponents of urban renewal and outlined several alternative hypotheses for sustaining successful urban environments. Many of her ideas were based on her own anecdotal observations of urban residents, but can now be investigated quantitatively using recently available urban data.

Jacobs was a pioneering voice for the concept of urban *vibrancy*, a measure of positive activity or energy in a neighborhood that makes an urban place unique and enjoyable to its residents despite the challenges of urban living. An important term coined by Jane Jacobs was "eyes on the street" which summarized her viewpoint that safer and more vibrant neighborhoods were those that had many people engaging in activities (either commercial or residential) on the street level at different times of the day (Jacobs 1961).

This concept of eyes on the street has been more recently expressed as the "natural surveillance" component of the *Crime Prevention through Environmental Design* movement (Deutsch 2016). These contemporary theories argue that the likelihood of criminal activity is strongly linked to the presence or absence of people on the street. As Deutsch (2016) states: "Criminals do not like to be seen or recognized, so they will choose situations where they can hide and easily escape." Policies which promote vibrancy and activity could potentially benefit crime prevention.

The recent explosion in high resolution data on cities offers an exciting opportunity for quantitative evaluation of contrasting urban development perspectives as well as current urban planning efforts. In this paper, we outline a pipeline for data collection and analysis of the associations between neighborhood safety, economic and demographic conditions and the built environment within urban environments.

We target our analysis pipeline towards a more specific goal: using high resolution data to create quantitative measures of the built environment that can serve as proxies for the human *vibrancy* of a local area. We then investigate the association between these vibrancy measures and safety in the city of Philadelphia. We focus on vibrancy proxy measures based on land use as well as business activity, which follows the "natural surveillance" idea that the presence of open businesses encourages safety through the store front presence of both staff and customers.

MacDonald (2015) provides a comprehensive review of past research into the association between the built environment and safety, where many quasi-experimental studies have shown that changes in housing, zoning and public transit can help to manage crime. In Section 5, we will try to emulate a quasi-experimental setting in our own analysis by comparing locations within census block groups, thereby matching locations in terms of economic health and population density.

The effects of natural surveillance on neighborhood vibrancy can be both subtle and complicated. The presence of a commercial business can encourage vibrancy through the presence of many people coming and going, or can give a sense of vacancy and isolation to an area if it is closed during a particular time of the day. In order to get an accurate picture of whether commercial businesses help to encourage safety, we will need to examine whether or not those businesses are open and active, as we outline in Section 4.

We choose the city of Philadelphia as a case study for this work as Philadelphia is currently encountering many contemporary issues in urban revival, population growth and desirability. Recent work has shown that urban city centers are growing relative to their suburban counterparts in many areas of the country (Couture and Handbury 2015). Another study by Ellen et al. (2016) finds an association between population movement of high-income and college-educated households and declining crime rates in central city neighborhoods.

We first outline our data collection for the city of Philadelphia in Section 2 and then explore the associations between safety and several economic, population and land use measures in Section 3. To get a more detailed picture of neighborhood vibrancy, we compile a database and several measures of business vibrancy in Section 4. In Section 5, we employ several matching analyses to evaluate the association between business vibrancy and safety within local neighborhoods, and then conclude with a brief discussion in Section 6.

In order to encourage replication of our urban analyses and adaptation to other research questions, we have made the code and public data that were used in our analyses available as a github repository at: `https://github.com/ColmanHumphrey/urbananalytics`

2 Urban Data in Philadelphia

Our analysis will be based on the geographical units defined by the US Census Bureau. Philadelphia county is divided into 384 census tracts which are divided into 1,336 block groups which are divided into 18,872 blocks. Figure 1 (left) gives a map outlining the 1,336 block groups in Philadelphia. Population and economic data are provided by the US Census Bureau, crime data is provided by the Philadelphia Police Department and land use data is provided by the City of Philadelphia.

A general theme of our urban work is that results can vary (often substantially) depending on the resolution level of the data and the geographic scale of the underlying processes involved. Most of our analyses will be done at the block group level which allows for the best interoperability between our economic and built environment data, but we also perform several analyses at the block level.

We use shape files provided by the US Census Bureau for our population and economic data and shape files provided by the City of Philadelphia for the land use data. Shape files from the Census Bureau delineate the boundaries and area of each census block and block group. Shape files from the City of Philadelphia delineate the boundaries and area of each lot in Philadelphia. For the vast majority of lots in Philadelphia, the lot is entirely contained within a single Census Bureau block. We outline further details of each data source below.

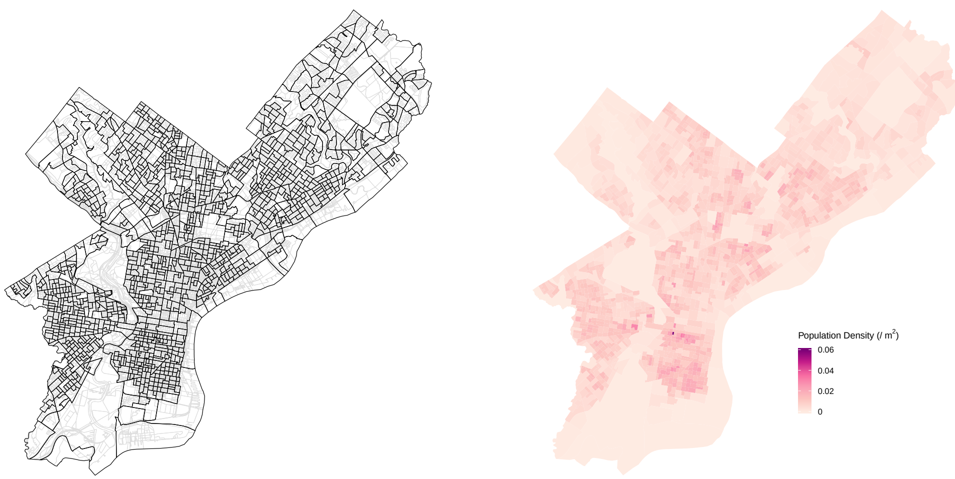

Figure 1. Left: Map of Philadelphia divided into block groups (black lines) by US Census Bureau. **Right:** Population density by block group in Philadelphia

2.1 Population Data

Our population demographic data were pulled from the census website (`factfinder.census.gov`) by setting the geography as all blocks in Philadelphia and setting the data source as "Hispanic or Latino Origin By Race" (which is SF1 P5 in their database). The raw demographic data gives the population count in each block from the 2010 census.

Of the 18,872 blocks in Philadelphia, 4,558 have no residents (e.g. parks, industrial areas, etc.). At the block level, we restrict our analysis to to blocks with at least 25 people, which gives 12,874 blocks that contain 98.9% of the population. At the block group level, we restrict our analysis to block groups with at least 400 people in them. which gives 1,325 block groups (out of 1,336) that contain 99.96% of the population.

We calculate the population count and population density in each block group i from the raw population data and using the area of each block group from the US Census Bureau shape files. Figure 1 (right) gives the spatial distribution of the population density across Philadelphia.

2.2 Economic Data

Our economic data were pulled from the American Community Survey on the census website (`factfinder.census.gov`): tables B19301 for income and C17002 for poverty, both from 2015. This data is only available at the block group resolution level.

For each block group in Philadelphia, we have the per-capita income and the fraction of the population in seven different brackets of income-to-poverty-line ratios: $[0, 0.5), [0.5, 1), [1, 1.25), [1.25, 1.5), [1.5, 1.85), [1.85, 2),$
$[2, \infty)$. For interpretation, the $[0.5, 1)$ bracket represents families that have income between 50% of the poverty line and the poverty line.

The poverty line threshold for each household is defined by the Census Bureau according to the size and composition of the household. As an example, a family of four with two children has a poverty line threshold of $23,999.

We define a scalar poverty measure for each block group based on the weighted sum of the proportion of block group households in each of the seven poverty brackets:

$$\text{Poverty}_i = \sum_{q=1}^{7} w_q \, p_{i,q}$$

where $p_{i,1}$ is the proportion of block group i households in the lowest bracket $[0, 0.5)$ and $p_{i,7}$ is the proportion of block group i households in the highest bracket $[2, \infty)$. We employ linearly decreasing weights $\mathbf{w} = [1, 5/6, 4/6, 3/6, 2/6, 1/6, 0]$ to give highest weight to the brackets with highest poverty. Our Poverty_i metric varies from 0 to 1, with larger values implying higher poverty: a block group with every household in the $[2, \infty)$ bracket takes the value zero, and one with every household in the $[0, 0.5)$ bracket takes the value one.

Figure 2 gives the spatial distribution of income Income_i (left) and our poverty metric Poverty_i (right) at the block group level in Philadelphia. We see that the areas of West Philadelphia and North Philadelphia have the lowest incomes and highest levels of poverty.

2.3 Land Use Zoning Data

Land use zoning data were downloaded from the City of Philadelphia. The land use data consists of a shapefile that divides the city into approximately 560,000 lots and provides the area and registered land use zoning designation (commercial, residential, industrial, vacant, transportation, water, park, civic, recreation, culture, and cemetery) for each of these lots. As an example, we show the land use designations for overall Philadelphia and the center city neighborhood in Figure 3.

Note that we combine the "commercial business" and "commercial consumer" into a single *commercial* designation, and all three "residential" densities into a single *residential* designation. For the rest of this paper, *Mixed Use* refers to the designation of "commercial / residential mixed".

We create several land use measures from these zoning designations. First, we calculate the fraction of area in each geographic unit (either block or block group) i that is designated as 'Vacant',

$$\text{Vacant.Prop}_i = \frac{Area_i(\text{Vacant})}{Area_i}$$

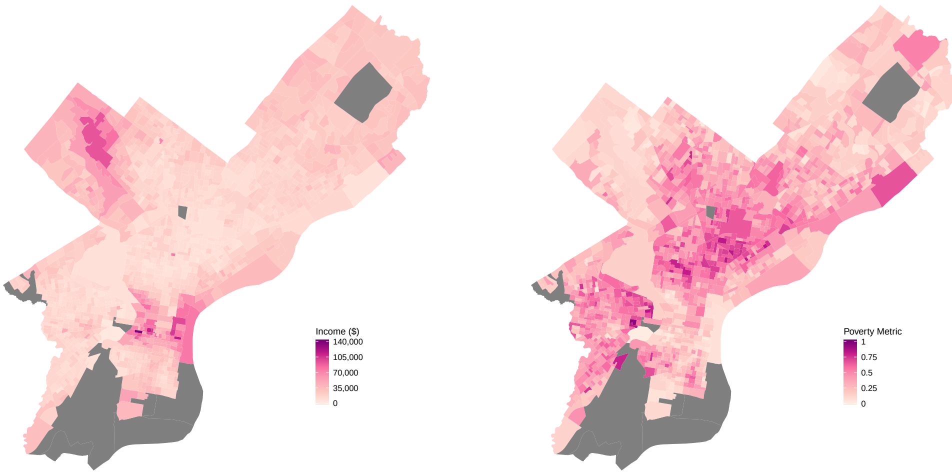

Figure 2. **Left:** Per-capita income **Right:** Poverty metric. Block groups that are colored grey do not have enough residents for the US Census Bureau to report economic data for those block groups.

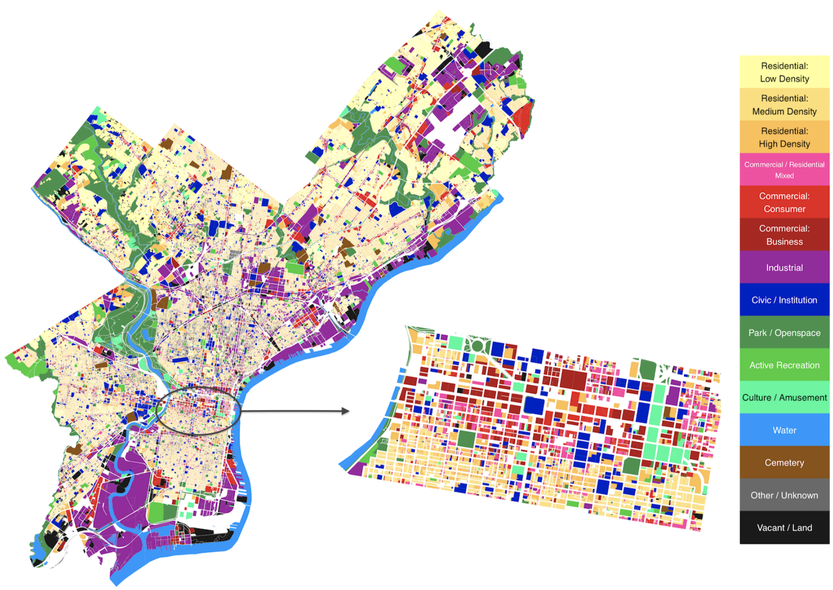

Figure 3. Land use designations for overall Philadelphia and the center city neighborhood

Second, we calculate the ratio of the area in each geographic unit (either block or block group) i that is commercial versus residential,

$$\text{ComRes.Prop}_i = \frac{Area_i(\text{Commercial})}{Area_i(\text{Commercial}) + Area_i(\text{Residential})}$$

Finally, we calculate a mixed use proportion, i.e. the proportion of every block or block group that is designated as mixed use,

$$\text{MixedUse.Prop}_i = \frac{Area_i(\text{Mixed Use})}{Area_i}$$

These land use zoning metrics provide our first set of proxy measures for the vibrancy of a local neighborhood. Figure 4 gives the spatial distribution of vacant proportion (left) and mixed use proportion (right) at the block group level in Philadelphia.

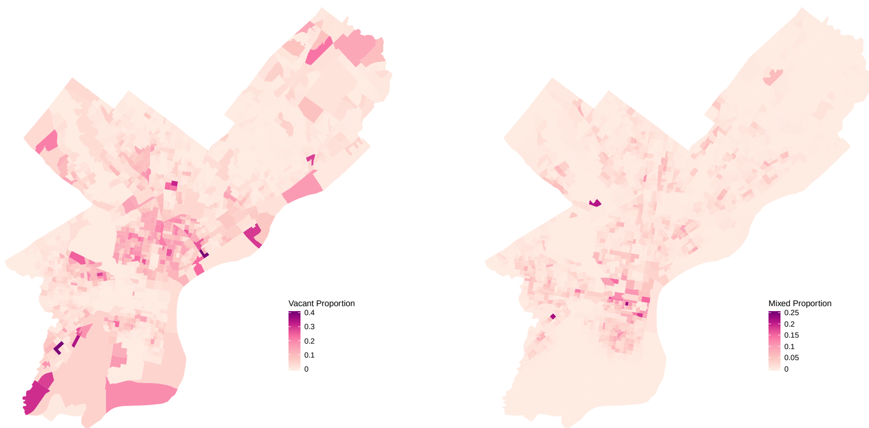

Figure 4. **Left:** Vacant Proportion **Right:** Mixed Use Proportion

Philadelphia's zoning procedures were revised in 2012 (`http://www.phila.gov/li/Pages/Zoning.aspx`). Our zoning data was downloaded in June 2014, and all of our analyses are based on that snapshot. While most of the city's zoning remains unchanged, lots can be rezoned through applications on a continuous basis.

2.4 Crime Data

Crime data for Philadelphia comes from the Philadelphia Police Department through the opendataphilly.org website. From their documentation: *Data comes directly from the Police Departments mainframe INCT system. The INCT system is fed by field incident reports and Computer Aided Dispatch system.* Our dataset consists of all crimes reported by the police in the city of Philadelphia from January 1, 2006 to December 31, 2015.

For each crime, we have the type of crime, the date and time of the crime, and the location of the crime in terms of latitude and longitude (WGS84 decimal degrees). Each crime in our dataset is categorized into one of several types that are listed along with the relative frequency of those types in Figure 5. Note that these crime categories are roughly ordered in terms of severity, and that high severity crimes are much less frequent.

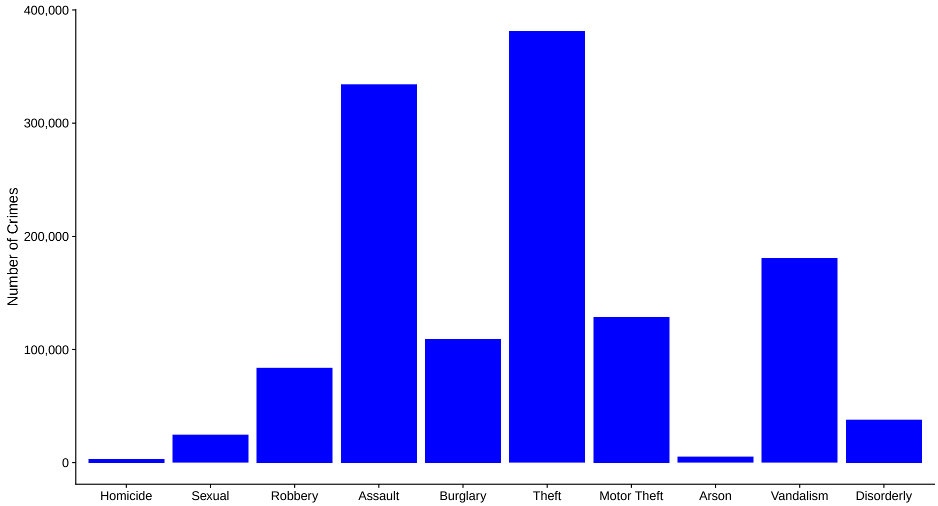

Figure 5. Relative frequency of different crime types reported in Philadelphia from January 1, 2006 to December 31, 2015.

For our subsequent analysis, we will combine these categories into two super-categories of crimes: a. **violent crimes** (Homicides, Sexual, Robbery and Assault) and b. **non-violent crimes** (Burglary, Theft, Motor Theft, Arson, Vandalism, and Disorderly Conduct).

Figure 6 gives the spatial distribution by block group of violent vs. all crimes committed in Philadelphia from 2006-2015. We see substantial heterogeneity in crime counts across the city with a large outlier count of both violent and non-violent crimes in the Market East block group of center city.

3 Exploring Neighborhood Factors Associated with Safety in Philadelphia

3.1 Association between Crime and Population

We first examine whether population Count_i and/or Density_i are associated with either violent or non-violent crimes in Philadelphia. Figure 7 plots the relationship between these two population measures and violent vs. non-violent crimes. Figure 7 also include the correlation and test statistic for the slope from a robust regression that downweights outlying values (Huber 1981). We also explored Poisson and Negative Binomial regressions but found that these alternative formulations did not give substantially different results.

We see in Figure 7 that population count is more strongly associated with both violent crime and non-violent crime than population density. In fact, population density is not significantly associated with violent crime, and negatively associated with non-violent crime.

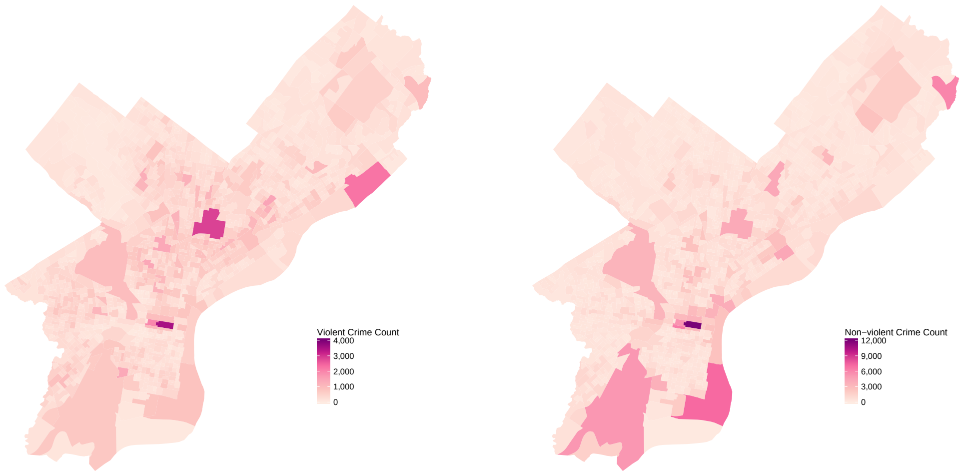

Figure 6. Number of Violent and Non-Violent Crimes in Philadelphia: 2006-2015

The lack of a strong positive association between population density and crime is especially notable in the context of popular historical hypotheses such as Simmel (1971) which argue that high population density leads to negative attitudes and hostility. In contrast, we find population size to be more strongly associated with crime compared to population density, which supports the work of Verbrugge and Taylor (1980).

In order to incorporate the association between crime and population count into our subsequent analyses, we define *excess violent crime* in each block group as the residual violent crime for that block group from the robust regression of violent crime on population count. Similarly, we define *excess non-violent crime* in each block group as the residual non-violent crime for that block group from the robust regression of non-violent crime on population count. In other words, the variables for safety in the next Section 3.2 will be excess crime (violent or non-violent) beyond the expected crime based on population count.

3.2 Association between Excess Crime and Economic Measures

As outlined in Section 2.2, we focus on two measures of the economic health of each block group in Philadelphia: 1. per-capita income and 2. our poverty metric. Figure 8 plots the relationship between these two economic measures and excess violent versus non-violent crime.

In Figure 8, we see a strong negative relationship between excess violent crime and income ($r = -0.44$) and a strong positive relationship between excess violent crime and poverty ($r = 0.59$). There is also noticeable non-linearity in the relationship between income and violent crime, with an even stronger linear relationship between violent crime and income for per-capita income below \$50,000 but much less of a relationship above per-capita income of \$50,000.

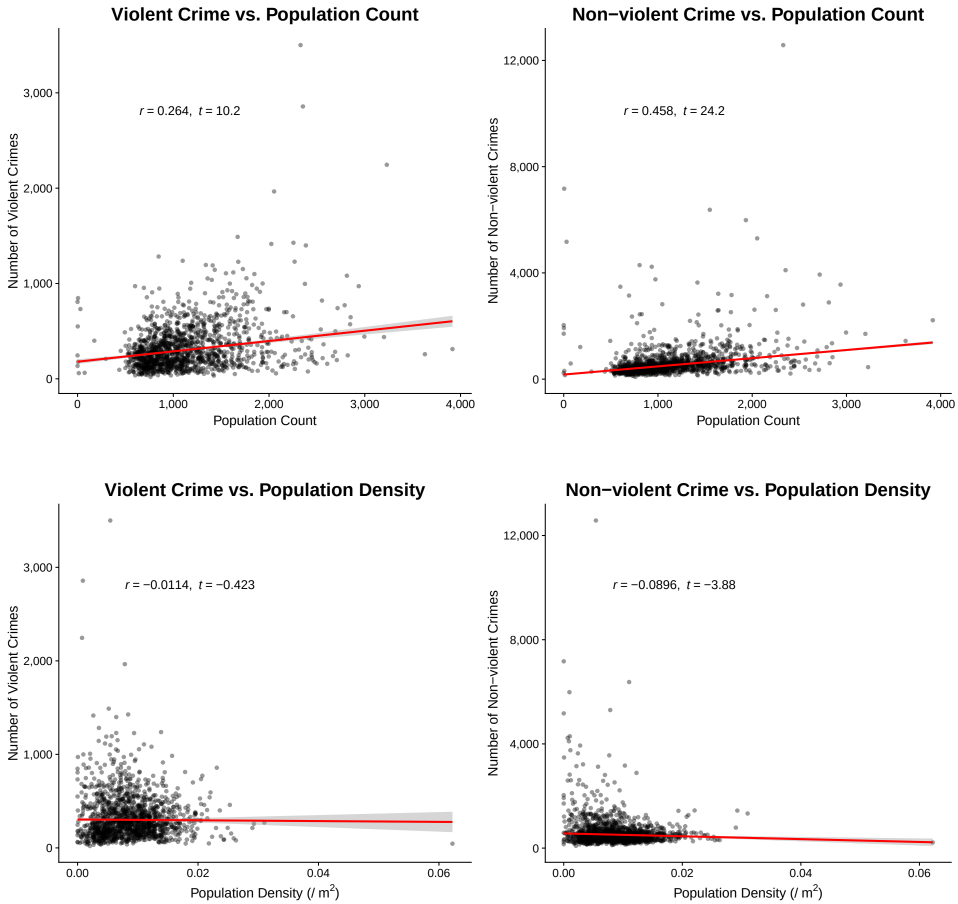

Figure 7. Association between Safety and Population. Predictor variables are either population count (top row) or population density (bottom row). Outcome variables are either violent crime counts (left column) or non-violent crime counts (right column). Red lines (and grey bands) are the least squares lines (and confidence bands) from a robust regression that downweights outlying values.

These economic measures have a much weaker relationship with excess non-violent crime. There is a weak negative association between per-capita income and excess non-violent crime ($r = -0.12$) and a weak positive association between poverty and excess non-violent crime ($r = 0.33$). Together these results suggest that per-capita income and poverty are strongly associated with excess violent crime but not excess non-violent crime, possibly because non-violent crimes are more crimes of opportunity occurring in areas located away from where the perpetrators of those crimes reside. Crimes of opportunity may be more driven by locations of businesses (rather than residences) which helps to motivate our work in Sections 4-5.

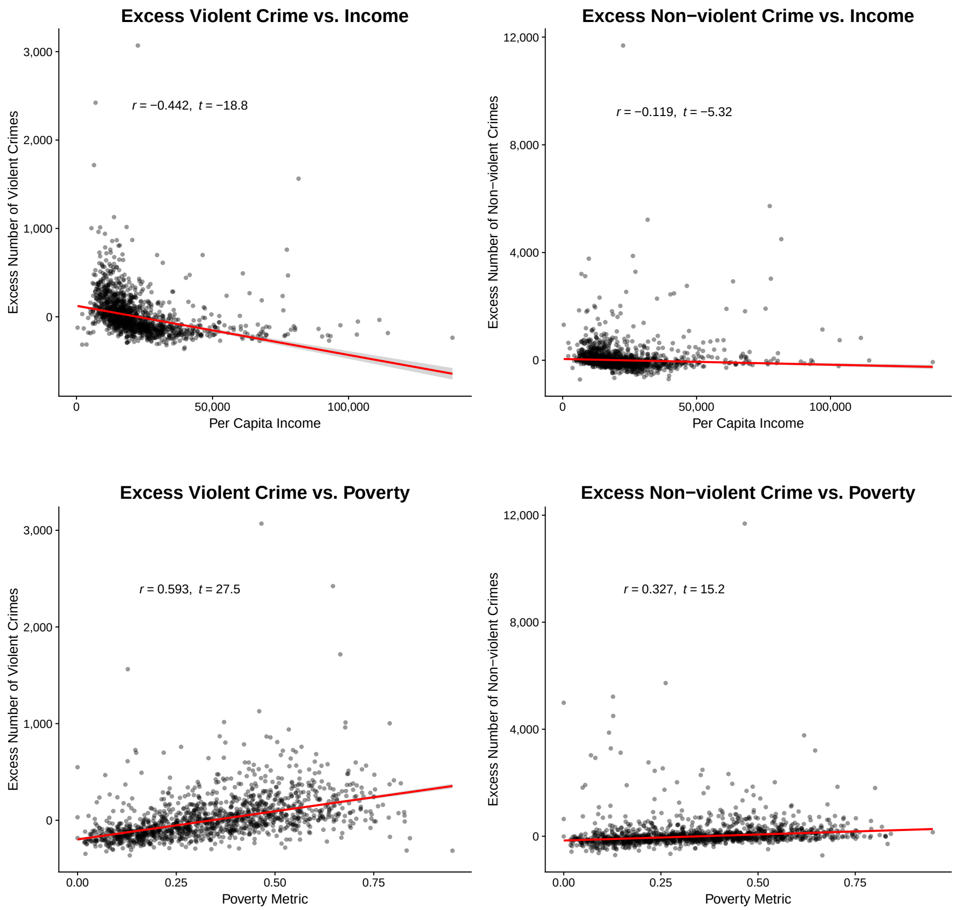

Figure 8. Association between safety and economic measures. Predictor variables are either per-capita income (top row) or our poverty metric (bottom row). Outcome variables are either excess violent crime counts (left column) or excess non-violent crime counts (right column). Red lines (and grey bands) are the least squares lines (and confidence bands) from a robust regression that downweights outlying values.

In order to incorporate the association between crime and these economic measures into our subsequent analysis, we now re-define *excess violent crime* in each block group as the residual violent crime in that block group from the robust regression of violent crime on population count, per-capita income and our poverty metric; Similarly for *excess non-violent crime*. So for the next Section 3.3, excess crime (violent or non-violent) in a block group is the number of crimes beyond expectation based on population count, income and poverty.

3.3 Association between Excess Crime and Land Use Zoning

Up to this point in our exploratory data analyses, we have focussed on the relationship between safety and features based on residents, namely the population and economic health, of each neighborhood. However, our primary goal is to investigate the role that the *built environment* of the neighborhood plays in safety, since effects of the built environment could inform future public policy initiatives.

As presented in Section 2.3, one type of data that we have pertaining to the built environment is the land use zoning designations for each lot in the city of Philadelphia. We used those zoning designations to create three measures of the "vibrancy" in each block group i: the fraction of vacant land (Vacant.Prop_i), the fraction of mixed use land (MixedUse.Prop_i) and the ratio of commercial area to residential area (ComRes.Prop_i). Figure 9 plots the relationship between these three land use vibrancy measures and excess violent versus excess non-violent crime.

Examining Figure 9, we see a moderately strong positive relationship between vacant proportion and excess violent crime ($r = 0.2$) and a similar relationship between vacant proportion and excess non-violent crime ($r = 0.20$). We see a stronger positive relationship between commercial vs. residential proportion and excess violent crime ($r = 0.42$) and an even stronger positive relationship between commercial vs. residential proportion and excess non-violent crime ($r = 0.65$). Finally, we see moderately strong positive relationship between mixed use proportion and excess violent crime ($r = 0.23$) and between mixed use proportion and non-violent crime ($r = 0.23$).

The moderately strong positive relationship we find between vacant lots and safety is related to recent investigations into the effect of "greening" of vacant lots on neighborhood safety (Branas et al. 2011). In that study, vacant lots that were randomly selected to be turned into green spaces were compared with a control set of vacant lots without an intervention. Branas et al. (2011) found that the "greening" of vacant lots was associated with a reduction of certain crime types and promotion of some positive health outcomes.

The strong positive relationship we find in Figure 9 between commercial proportion and crime is also very interesting in the context of contemporary theories of urbanism. As we describe in Section 1, the "eyes on the street" theory of Jacobs (1961) and "natural surveillance" theory of (Deutsch 2016) argue that safer and more vibrant neighborhoods were those that have greater presence of people on the street achieved through a mixing of commercial and residential properties.

Our findings in Figure 9 do not support the idea that a mix of commercial and residential land use leads to increased safety. However, we must concede that land use zoning designations are a rather imperfect and low resolution indication of urban vibrancy. Land use zoning demonstrates a type of use deemed appropriate based on proximity to residents as well as governing fire safety and exposure. Zoning also regulates building heights and the type of allowable activity in a given structure. In particular, the zoning designation of a particular lot as commercial does not provide insight into whether the commercial enterprise located in that lot contributes positively to vibrancy of the area.

This type of data also does not contain information about whether that commercial enterprise is open or closed during times when crimes tend to be committed. As we describe in Section 1, an indication of the activity of a commercial business is important to evaluating its impact on the neighborhood. An open business can encourage vibrancy through the activity of its staff and customers, or can give a sense of vacancy and isolation to an area if it is closed.

Thus, key information for urban vibrancy is missing from the land use zoning data, such as the types of business occupied on commercial property and when those businesses are open. This missing information

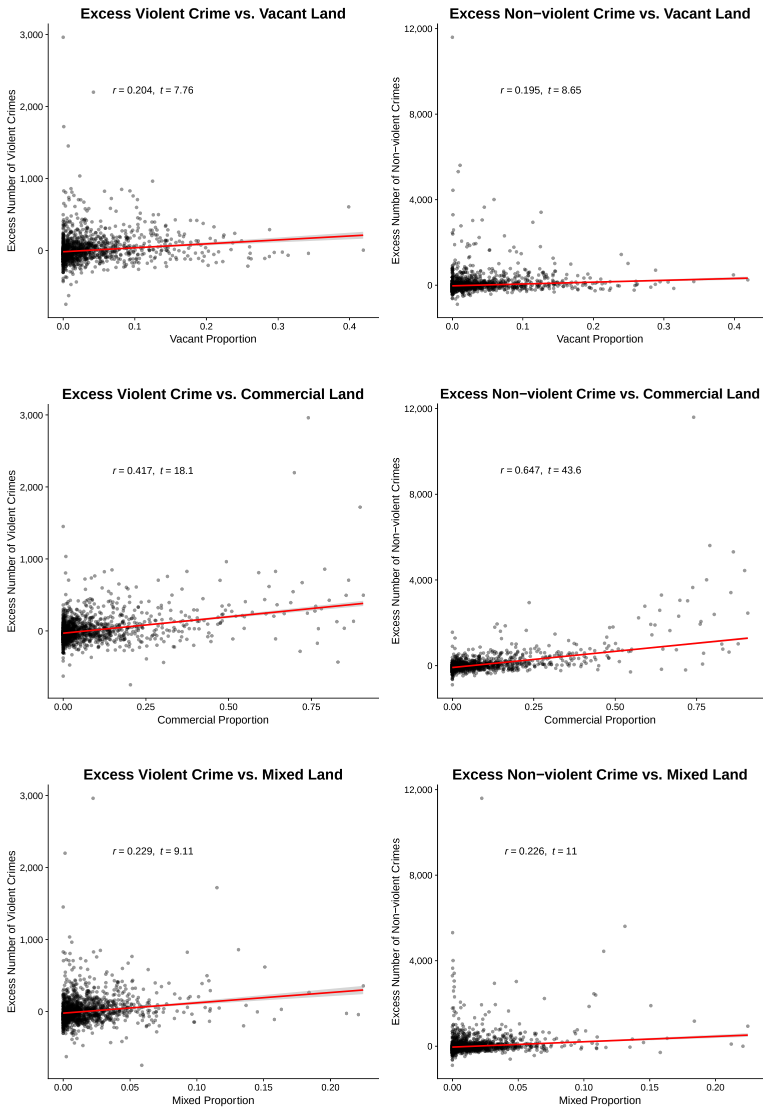

Figure 9. Association between safety and land use vibrancy measures. Red lines (and grey bands) are the least squares lines (and confidence bands) from a robust regression that downweights outlying values.

motivates our investigation of more detailed measures of neighborhood vibrancy based on business data in the following Section 4.

4 Urban Vibrancy Measures based on Business Data

As discussed in Section 3.3, measures based on land use zoning designations are an insufficient summary of the vibrancy of a neighborhood. We can not evaluate whether a mix of commercial and residential properties promotes safety without first establishing what types of business enterprises are present in lots zoned for commercial use. We need to better understand when businesses are active, what type of business they are, and how they contribute to vibrancy. To that end, we outline our manual assembly and curation of a database of Philadelphia businesses, as well as the construction of several measures of business vibrancy from that data.

4.1 Business Data

We have manually assembled a database of Philadelphia businesses by scraping three different web sources: Google Places, Yelp, and Foursquare. Each of these sources provide the GPS locations for a large number of businesses in Philadelphia, as well as opening hours for a subset of those businesses.

The most difficult issues with assembling this business database are: 1. integrating these three data sources and removing overlapping businesses and 2. categorizing all businesses into a small set of *business types*. Table 1 gives the number of businesses and the number of those businesses where we have opening hour information. We also give counts of the total number of businesses and the number of businesses with opening hours in the union of all three data sources (removing duplicates between data sources).

Table 1. Number of businesses and number with opening hours from each data source.

	Google	Yelp	Foursquare	Union
Total businesses	$34,768$	$12,534$	$40,331$	$72,020$
Businesses with hours	$12,346$	$7,728$	$7,022$	$19,140$

Each data source has its own categorization for businesses, with Google using about a hundred categories and Yelp and Foursquare each using closer to a thousand categories. Out of the myriad of business categories across all data sources, we created ten *business types*: Cafe $(4,166)$, Convenience $(1,481)$, Gym $(1,273)$, Institution $(24,489)$, Liquor (316), Lodging (461), Nightlife $(5,108)$, Pharmacy (799), Restaurant $(7,909)$, and Retail $(31,147)$. The values in parentheses are the total number of businesses in each business type.

A particular business can belong to multiple business types, e.g. a restaurant that also sells liquor to go. Most of these business types are self-explanatory, but we need to clarify a few details. The cafe type includes cafes, bakeries and coffee shops that are not full restaurants. The restaurant type also includes meal delivery and meal take out businesses. Institution is a broad type that includes banks, post offices, churches, museums, schools, police and fire departments, as well as many others.

4.2 Measures of Business Vibrancy

We use our assembled business data to create several high resolution measures of business vibrancy at any particular location in the city of Philadelphia. We want these measures to encapsulate whether a given location has a concentration of a businesses of a particular type, and whether those businesses are active storefronts (i.e. open) during times of the week when crimes tend to be highest.

We examined the frequency of crimes at different times of the week and isolated two "high crime windows" that have a disproportionately large number of crimes (both violent and non-violent) relative to other times of the week. These two high crime windows are *weekday evenings*, which we define as 6-12pm Monday-Friday and *weekend nights*, which we define as 12-4am Saturday-Sunday.

In Section 5, we will evaluate whether business vibrancy is associated with crime totals during these two specific high crime windows as well as throughout the entire week.

The first set of measures of business vibrancy we consider are simply the total number of businesses of each *business type* near to any particular location in the city. Recall that our ten business types are Cafe, Convenience, Gym, Institution, Liquor, Lodging, Nightlife, Pharmacy, Restaurant, and Retail. We expect that some of these business types will be more associated with safety than others.

In addition to the total number of businesses of each type near to a particular location, we want to take into account whether those businesses are active storefronts in the sense of being open. In particular, we are interested in whether a given location has businesses of a particular type (e.g. cafes) that are open longer than expected.

We first establish a *consensus* number of open hours for each business type by calculating the average open hours across all businesses of that type in Philadelphia. We can then calculate the *excess* number of open hours for each business (for which we have open hours) in Philadelphia relative to the consensus for that business type. For example, a particular cafe will have *excess* of 2 if it is open for 2 hours more than the consensus open hours for all cafes in Philadelphia, whereas a particular restaurant will have *excess* of -3 if it is open for 3 hours less than the consensus open hours for all restaurants in Philadelphia.

Building upon these calculations, the second set of measures of business vibrancy we consider are the average excess hours of businesses of each *business type* near to any particular location in the city. We calculate these excess hour measures over the entire week as well as just within in the two high crime windows (weekday evenings and weekend nights).

In summary, we have two sets of measures of the business vibrancy around any particular location: the number of businesses of each business type and the average excess hours of each business type. The latter can be calculated over the entire week or just within the high crime windows mentioned above.

In Section 5, we evaluate the association between these business vibrancy measures and both excess violent and non-violent crimes within the local neighborhoods defined by our census block groups.

5 Evaluating the Association between Business Vibrancy and Safety

With our new business vibrancy measures in hand, the goal of our analysis is evaluating the association between these measures and safety at the local neighborhood level, while controlling for the characteristics of those neighborhoods.

We will control for neighborhood characteristics by focusing our analyses on comparing pairs of locations within each block or within each block group. Underlying this strategy is an assumption that the census blocks or block groups are small enough that different locations within these areal units (blocks or block groups) should be highly similar with regards to the demographic and economic measures we examined in Sections 3.1 and 3.2.

We explore two types of within-block comparisons. In Section 5.1, we find pairs of locations within block groups where one location has businesses that are "open longer" relative to the consensus for their business type and where the other location has businesses that are "open shorter" relative to the consensus for their business type. We then examine these within-block-group pairs to see if there are differences in crime between "open shorter" vs. "open longer" locations.

In Section 5.2, we find pairs of locations within blocks where one location has the highest density of crimes and the other location has the lowest density of crimes within that block. We then examine these within-block pairs to see if there are differences our business vibrancy measures between the "high crime" vs. "low crime" locations.

5.1 Comparing "Open Shorter" vs. "Open Longer" Locations

For each of our ten business types, we identify block groups that contain a pair of businesses (of that type) where one of those businesses has long opening hours and the other business has short opening hours. We define a business as having long opening hours if its total opening hours are above the 75th percentile for businesses of that type. Similarly, we define a business as having short opening hours if its total opening hours are below the 25th percentile for businesses of that type.

We further restrict ourselves to block groups where the pair of businesses are at least 140 meters apart, which is roughly the size of a Philadelphia city block. It should be noted that only a small subset of the 1,336 block groups in Philadelphia will contain such a valid pair of businesses: both a long opening and short opening business of a particular type separated by at least 140 meters. As an example, lodging had only one block group in the entire city with a valid within-block group comparison for the total week comparison and so this business type is excluded from this analysis.

For each block group containing such a valid comparison, we then count the number of crimes that occurred within a 70 meter radius around both the long opening hour business and the short opening hour business (which ensures that we do not double count any crimes for both businesses).

The object of this analysis is the difference in crimes between the short opening hour business and the long opening hour business within each block group that contains such a business pair. If businesses that are active (open for a longer period) help to deter crime and promote safety, then these differences in crime should be positive. For each business type, we calculate a matched pairs mean differences in crime around the short opening hour minus long opening hour businesses in each within-block group pair.

In Figure 10, we display the matched pair mean differences in crime between short opening and long opening hour businesses of each business type separately. We calculate different matched pair mean differences for only violent crimes, only non-violent crimes and all crimes. The significance threshold for these t-statistics was Bonferroni-adjusted to account for the number of comparisons being tested. We also divide up these comparisons into the three different time windows discussed in Section 4.2: the entire week plus two high crime windows, *weekday evenings* and *weekend nights*.

Examining Figure 10, we see mostly negative differences (red) which imply that more crimes are occurring around the business location with longer open hours, especially nightlife locations and restaurants. A notable exception are gyms, which show positive differences which imply fewer crimes occurring around the gym with longer open hours, for all windows and both crime types.

Violent crimes near convenience locations are also an interesting case. Over the entire week, fewer non-violent crimes occurred around the convenience locations with longer open hours (blue), while approximately the same number of violent crimes occurred. For the weekday evening window, far fewer crimes of both types occurred near convenience locations with longer open hours (blue), but this trend is reversed during weekend nights, where more crimes occurred near convenience locations with longer open hours.

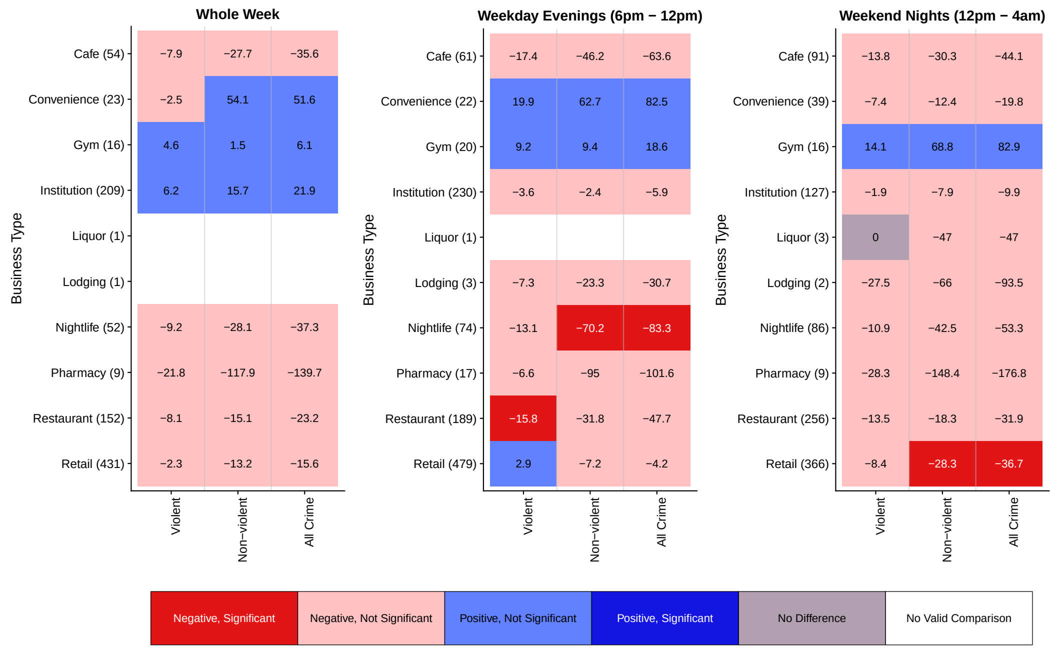

Figure 10. Matched pair mean differences in crime between short opening and long opening hour businesses, calculated separately for each combination of crime type and business type. Different panels are used to display the mean differences calculated over the entire week vs. just weekday evenings vs. just weekend nights. The significance threshold of $p = 0.05$ was Bonferroni-adjusted to account for multiple comparisons. Values in parentheses are the number of block groups with valid comparisons for that business type

Recall that our preliminary hypothesis, motivated by Jacobs (1961) and Deutsch (2016), was that greater business vibrancy would be associated with fewer crimes around those vibrant locations relative to less vibrant locations in the same block group. The results in Figure 10 for gyms does show a trend in this expected direction, but the results for several other business types goes against that hypothesis.

That said, there are not many differences in Figure 10 that are statistically significant. To a large extent, the lack of significance is driven by the small sample sizes in these comparisons. For example, there are only nine block groups with a pair of open shorter vs. open longer pharmacies for the whole week comparison, which does not give us much power to detect differences in crime associated with differences in business vibrancy.

Another weakness of this analysis is that we picked our locations for these comparisons based on a single "longer" open business and a single "shorter" open business. To incorporate a greater number of businesses in our comparisons, we can instead focus on comparing locations based on high versus low crime in the next Section 5.2.

5.2 Comparing “High Crime” vs. “Low Crime” Locations

In this comparison, we first calculate the location with the highest crime frequency and the location with the lowest crime frequency within each block. We perform this analysis on the census block level (rather than the block group level) in order to give an even higher resolution view of the association between vibrancy and safety. For each business type separately, we then calculate our measures of business vibrancy from Section 4.2 around both the high crime and low crime locations within each block.

Many blocks do not contain any businesses of particular business types near either high or low crime locations, which excludes those blocks from any comparisons involving that particular business type. We further restrict ourselves to blocks where the highest crime and lowest crime locations are at least 100 meters apart. Similar to Section 5.1, these restrictions limit the sample size for each of our comparisons.

For each block containing such a valid comparison, we calculate our two measures of business vibrancy, the number of businesses of each business type and the average excess hours of each business type, around the high crime and low crime locations in those blocks. For each business type, we calculate a matched pairs t-statistic for differences in the business vibrancy measures around the low crime location minus the the business vibrancy measures around high crime location within each block. If business vibrancy helps to deter crime and promote safety, then these differences in business vibrancy should be positive.

In Figure 11, we display the matched mean differences in the two business vibrancy measures (the number of businesses of each business type and the average excess hours of each business type) between the low crime and high crime within-block locations. We calculate differences for locations based on violent crimes and locations based on non-violent crimes. The significance threshold for these t-statistics was Bonferroni-adjusted to account for the number of comparisons being performed. We again also divide up these comparisons into the three difference time windows discussed in Section 4.2: the entire week plus two high crime windows, *weekday evenings* and *weekend nights*.

We see in Figure 11 that the number of businesses difference is significantly negative (red) for both violent and non-violent crimes for essentially all business types, most strongly retail stores and restaurants. This result suggests that there are more businesses around the higher crime locations than the lower crime locations.

However, we also observe in Figure 11 that for many of these business types, there are positive differences (blue) for our average excess hours metric, which implies that those businesses are open longer around the low crime location compared to the high crime location. These differences are not as significant, but we still see evidence of an interesting and subtle finding: more crimes tend to occur near business locations but fewer crimes tend to occur near businesses that are open longer, for cafes, and gyms. Note that the left hand plot, Whole Week, contains the largest comparison in terms of crimes and hours counted.

We can also compare our original land use zoning measures of vibrancy from Section 3.3 between these high and low crime locations. We again calculate differences for locations based on violent crimes and locations based on non-violent crimes, but now the differences are based on our three land use vibrancy measures: the fraction of vacant land, the fraction of mixed use land and the ratio of commercial area to residential area.

In Figure 12, we display the matched mean differences in the three land use vibrancy measures between the low crime and high crime within-block group locations. We again also divide up these comparisons

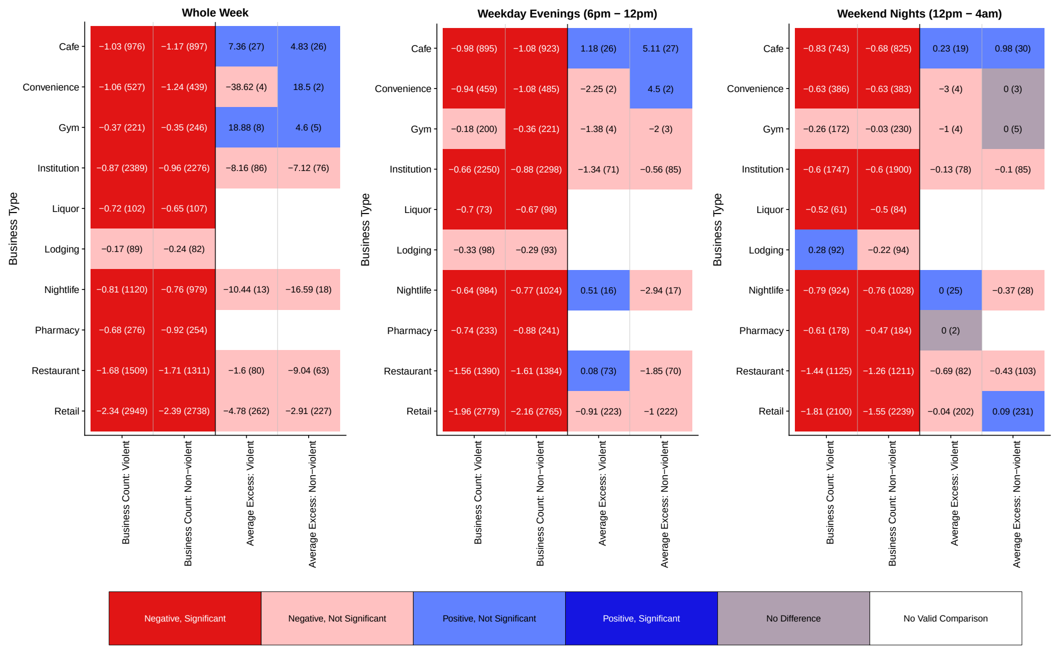

Figure 11. Matched pair mean differences in measures of business vibrancy between high crime and low crime locations, calculated separately for each combination of crime type and business type. Different panels are used to display the mean differences calculated over the entire week vs. just weekday evenings vs. just weekend nights. The significance threshold of $p = 0.05$ was Bonferroni-adjusted to account for multiple comparisons. Values in parentheses are the number of block with valid comparisons for that business type.

into the three time windows discussed in Section 4.2: the entire week plus two high crime windows, *weekday evenings* and *weekend nights*.

In Figure 12, we see very strong negative differences for mixed proportion and commercial vs. residential proportion, both of which strongly suggests that there is more mixed zoning and zoning for commercial use near to the high crime locations. This association between commercial enterprise and safety was also observed in Section 3.3 and motivated our development of more detailed measures of business vibrancy in Section 4.

We also see very strong positive differences for the vacant land proportion which suggests the presence of more vacant land near to low crime locations compared to the high crime locations. This finding is notable when compared to the positive association between vacant proportion and crime seen in Figure 9.

Together, those two findings suggest that neighborhoods with more vacant properties overall have higher crime but when looking within those neighborhoods, crimes tend not to be located near vacant properties. These results are especially interesting given the mixed effects on crime resulting from the "greening" of vacant lots in the study by Branas et al. (2011).

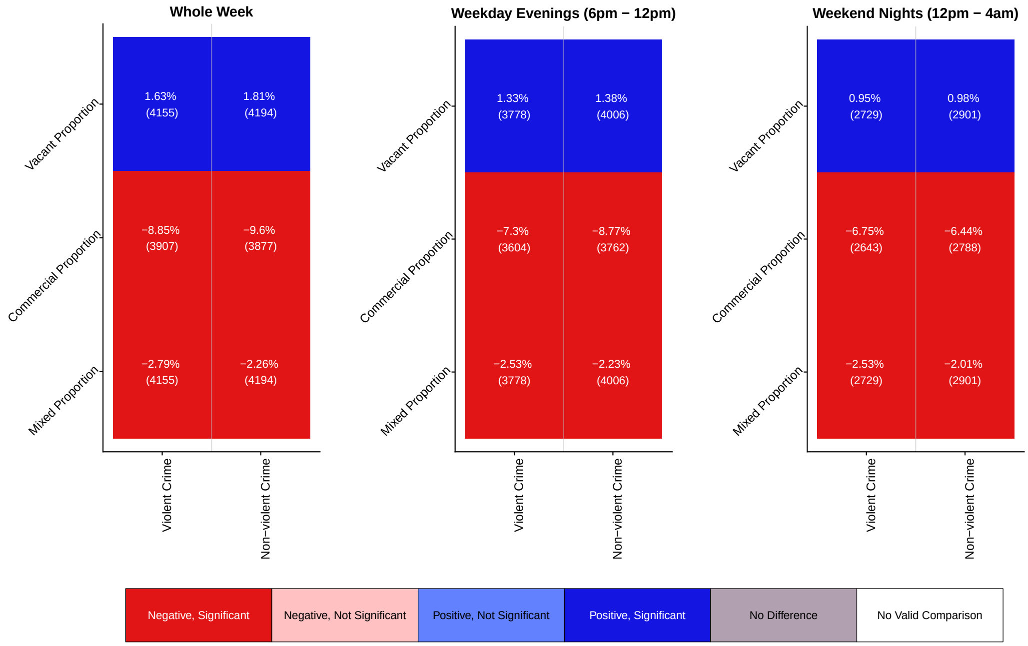

Figure 12. Matched pair mean differences in measures of land use zoning vibrancy, the fraction of vacant land and the ratio of commercial area to residential area, between high crime and low crime locations (calculated separately for each crime type and business type). Different panels are used to display the mean differences calculated over the entire week vs. just weekday evenings vs. just weekend nights. The significance threshold of $p = 0.05$ was Bonferroni-adjusted to account for multiple comparisons. Values in parentheses are the number of block groups with valid comparisons for that business type.

5.3 Summary of Business Vibrancy and Safety

Our analysis pipeline for studying the association between business vibrancy and safety has produced several findings that could impact current evaluations of contemporary theories in urban planning. First, we find that more crimes occur near business locations but that businesses that are open for longer periods are associated with fewer crimes. Second, we find that although neighborhoods with more aggregate vacancy have higher crime (Section 3.3), when comparing locations within each neighborhood, crimes tend not to be located near vacant properties.

Another important observation from Figures 10 and 11 is the substantial heterogeneity in the association between business vibrancy and crime both across different business types and different time windows. The power of both studies was compromised by small sample sizes as there are only a limited number of block groups that permit a pair of comparable locations. The associations between land use zoning and safety in Figure 12 are more significant due to much larger sample sizes of locations for these comparisons.

Clearly, the associations between safety and neighborhood vibrancy are subtle, heterogeneous, and in need of even higher resolution studies to fully understand. In Section 6, we discuss alternative strategies for matched comparisons that may permit more high resolution (and large sample size) analyses.

6 Discussion

The recent availability of high resolution data on cities provides a tremendous opportunity for sophisticated quantitative evaluation of historical and current urban development. To aid in this effort, we outline a framework for data collection and analysis of the associations between safety, economic and demographic conditions and the built environment within local neighborhoods. We used this framework to investigate a more specific goal: the creation of quantitative measures of "vibrancy" based on the built environment of a neighborhood and exploration of the association between these vibrancy measures and neighborhood safety.

We find that population density is not strongly associated with either violent or non-violent crime, which argues against the theory of Simmel (1971). We find that population count is a more important predictor of crime, which supports the work of Verbrugge and Taylor (1980). We also explored the association between crime and economic measures as well as measures of vibrancy derived from land use zoning data, but found that these measures were not an adequate summary of the local commercial vibrancy of an area.

To address vibrancy at a higher resolution, we constructed several measures of business vibrancy and employed matching of locations within block groups to evaluate the relationship between business vibrancy and safety. Our business vibrancy measures (number of businesses and average excess hours of businesses) are designed to be proxies for the "eyes on the street" concept of Jacobs (1961).

Our results suggest that more crimes occur near business locations but that businesses of some types that are active (open) for longer periods could be associated with fewer crimes. We also found that the overall proportion of vacancy in a neighborhood is associated with higher crime but that within a neighborhood, crimes tend to not occur near to vacant properties.

We also found substantial heterogeneity in the direction and strength of the association between crime and business vibrancy across different business types and different times of the week. Our view of the business vibrancy in a local area could be possibly improved by incorporating additional information such as more direct measures of business activity (beyond being open or not) when that data is available. Another potential option is business ratings, which are a primary feature of one of our commercial data sources, Yelp.

It may also be possible to perform our matching analyses at a higher resolution level, such as individual streets, rather than just locations within the same block group which may have more power for detecting subtle relationships. For example, Weisburd (2015) focussed on the street segment as their geographical unit of analysis when studying the concentration of crime.

It should also be noted that our simple testing procedures in Section 5 do assume that crimes are realized independently. This assumption is tenuous when there are multiple crimes reported from the same incident or dependence within perpetrators for repeated crimes and between co-perpetrators. However, we do not believe that these dependencies have had a substantial effect on our comparisons.

Outside of the business vibrancy measures that are the focus of this paper, there are many alternative data sources that would help to further define the vibrancy of local urban areas. Home prices would be a valuable resource for modeling the desirability of a neighborhood, though extensive publicly available home pricing data remains a challenge.

The company *Walk Score* produces a composite measure of the walkability of a neighborhood but their measure does not include several important details (Goodyear 2012), such as the types of available businesses which we found to be relevant in Section 5. The direct measure of foot traffic at the neighborhood or street level would certainly improve measures of urban vibrancy but this data is also not currently publicly available.

We encourage the adaptation of our analysis pipeline to other research questions within the urban analytics community. The code and public data that were used in our analyses is available as a github repository at: `https://github.com/ColmanHumphrey/urbananalytics`

7 Acknowledgments

This research was supported by a grant from the Wharton Social Impact Initiative (WSII) at the University of Pennsylvania.

References

Branas CC, Cheney RA, MacDonald JM, Tam VW, Jackson TD and Have TRT (2011) A difference-in-differences analysis of health, safety, and greening vacant urban space. *American Journal of Epidemiology* 174: 1296–1306.

Couture V and Handbury J (2015) Urban revival in america, 2000 to 2010. Working Paper 800, Samuel Zell and Robert Lurie Real Estate Center.

Deutsch W (2016) Crime prevention through environmental design. *The Balance* URL https://www.thebalance.com/crime-prevention-through-environmental-design-394571.

Ellen I, Horn K and Reed D (2016) Has falling crime invited gentrification? NYU Furman Center Working Paper.

Goodyear S (2012) Walk score is great, but it still doesn't capture walk appeal. *The Atlantic Citylab* .

Huber PJ (1981) *Robust Statistics*. Wiley, New York.

Jacobs J (1961) *The Death and Life of Great American Cities*. Random House, New York.

MacDonald J (2015) Community design and crime: The impact of housing and the built environment. *Crime and Justice* 44(1): 333–383.

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Simmel G (1971) The metropolis and mental life. In: Levine D (ed.) *Simmel: On individuality and social forms*. (Originally published in 1903), p. 324.

Verbrugge LM and Taylor RB (1980) Consequences of population density and size. *Urban Affairs Quarterly* 16: 135–160.

Walker K (2018) *tidycensus: Load US Census Boundary and Attribute Data as 'tidyverse' and 'sf'-Ready Data Frames*. URL https://CRAN.R-project.org/package=tidycensus. R package version 0.4.1.

Weisburd D (2015) The law of crime concentration and the criminology of place. *Criminology* 53(2): 133–157.

Wickham H, Francois R, Henry L and Mller K (2017) *dplyr: A Grammar of Data Manipulation*. URL https://CRAN.R-project.org/package=dplyr. R package version 0.7.4.

Winsborough HH (1965) The social consequences of high population density. *Law and Contemporary Problems* 30: 120–126.